Principles of Machine Learning: Session 2

Methusalem Colloquium Mini-Course

S. Nõmm

Department of Software Science, Tallinn University of Technology

08.12.2022

Deep learning

1. Area of machine learning sometimes called ‘AI’
2. Neural networks that can learn abstract representations
3. Natural networks of specific structure (different layers are responsible for different operations) that allows to solve problems of speech recognition, machine vision, text translation etc.
4. Particularly the following algorithms are considered to belong to the deep learning family:
   - **Convolutional Neural Networks (CNN)**
   - **Long Shot Term Memory (LSTM) networks**
   - Recurrent Neural Networks* (RNN)
   - **Generative Adversarial Networks (GAN)**
   - **Self Organising Networks (SOM)**
   - Autoencoders
   - **Transformers**
   - Multilayer perceptrons
   - Deep Belief Networks (DBN)
   - Restricted Boltzman Machines
   - …

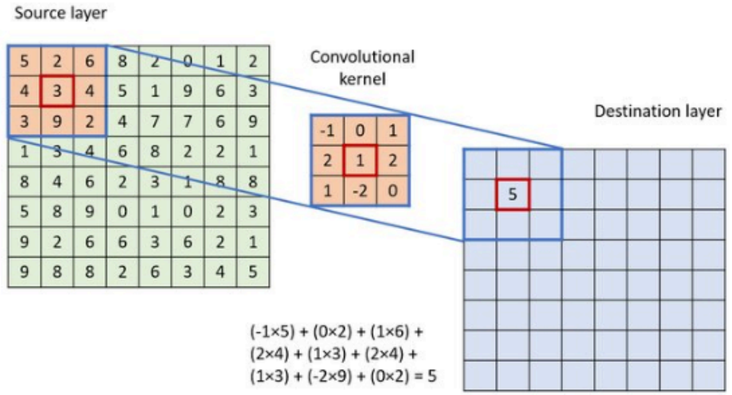
Deep learning: Convolutional Neural Networks (LeCun, 1989)

Convolution operation: $s(t) = (x * w)(t) = \sum_{a+-\infty}^{\infty} x(a)w(t-a)$.

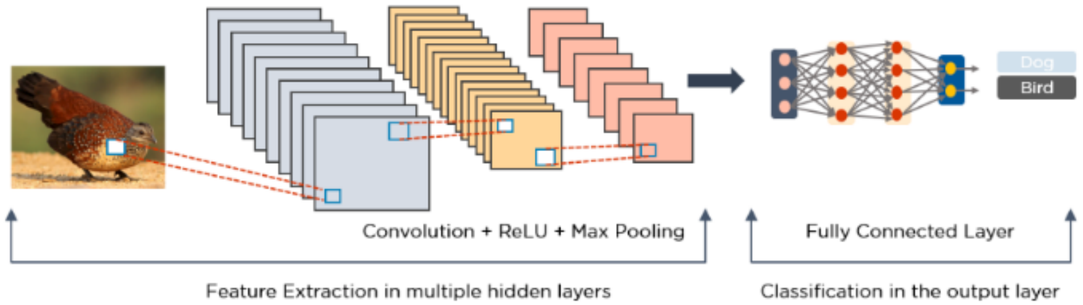
x - input, w - weighting function (referred as ***kernel***), s -output (referred as feature map).

CNN type networks are very popular in the area of image analysis where two-dimensional inputs and two dimensional kernels are in use

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n) .$$

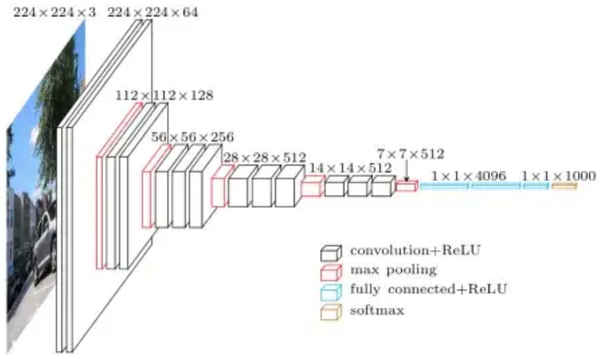


Deep learning: Convolutional Neural Networks



Typically CNN is structured as follows:

- Convolution layer
- Detector layer (nonlinear activation functions)
- Pooling layer (replaces outputs of the previous layer with the summary statistic of the neighbouring outputs) reduces dimensionality.
- Fully connected layer
- Soft max

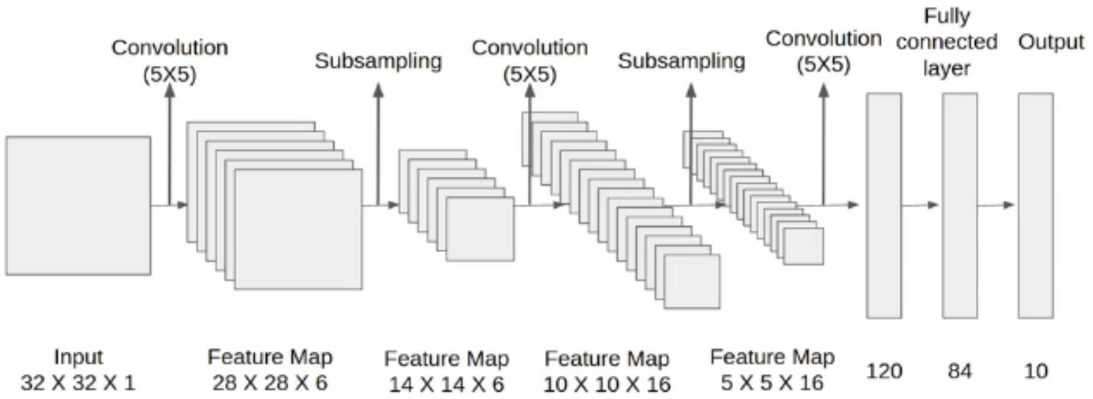


Architecture of VGG16 (a CNN model)

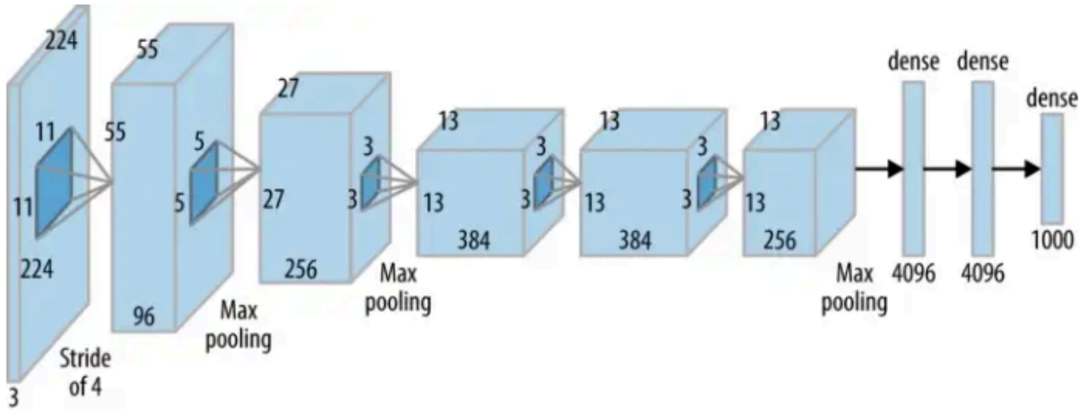
Softmax

$$\sigma(Z)_i = \frac{e^{z_i}}{\sum_j^K e^{z_j}}$$

Deep learning: CNN



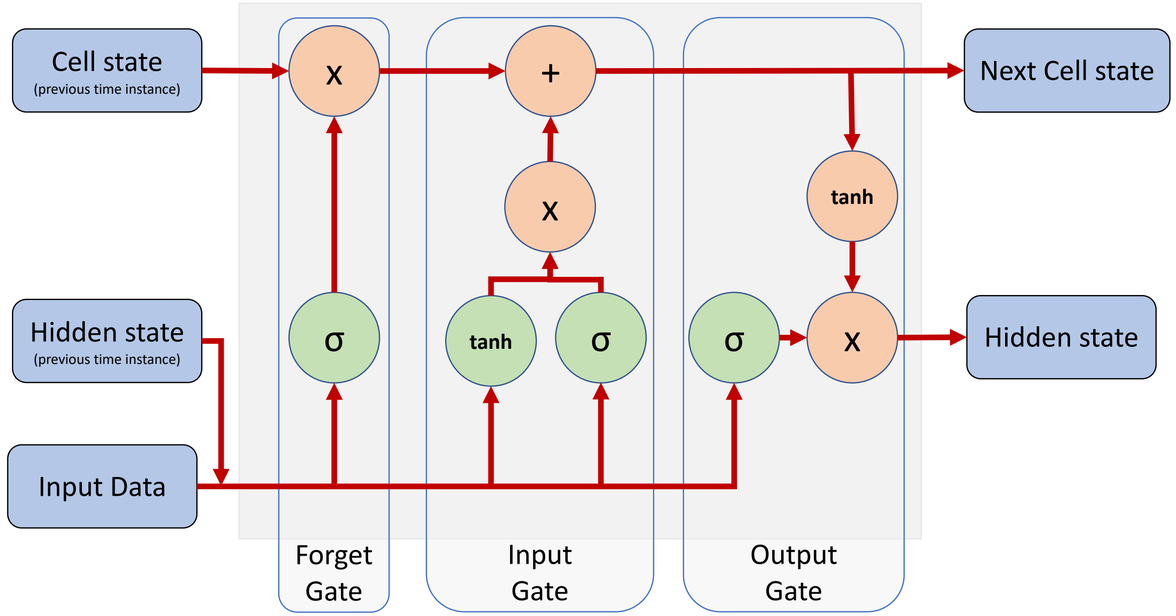
LeNet - 5 (LeCun, 1989)
60000 trainable parameters.

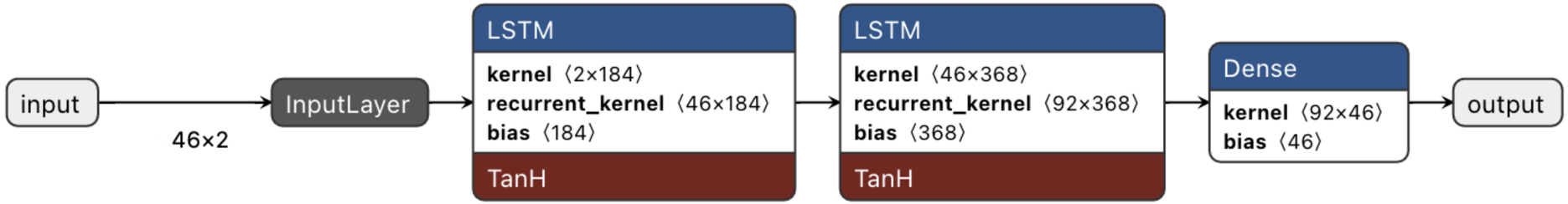


Alexnet Block Diagram (source:oreilly.com)

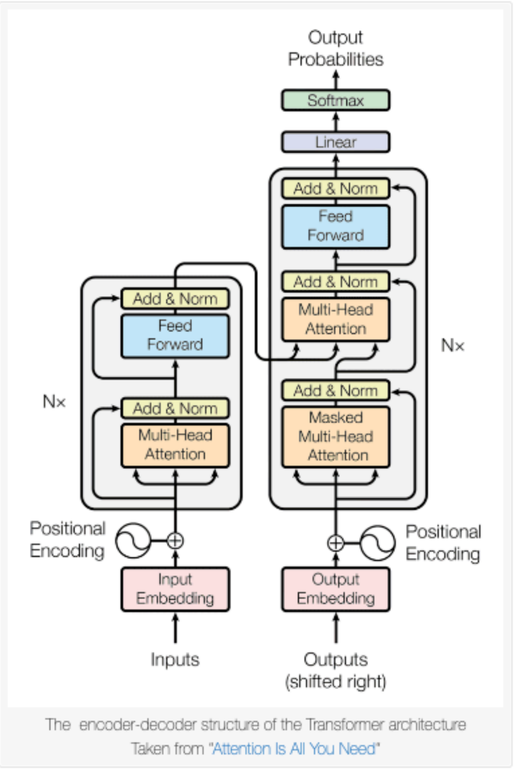
Nearly 63 000 000 trainable parameters.

Deep learning: Long Short Term Memory network (S.Hochreiter and J. Schmidhumber 1995)





Deep learning: Transformers (Vaswany et al. 2017)



The encoder-decoder structure of the Transformer architecture
Taken from "Attention Is All You Need"

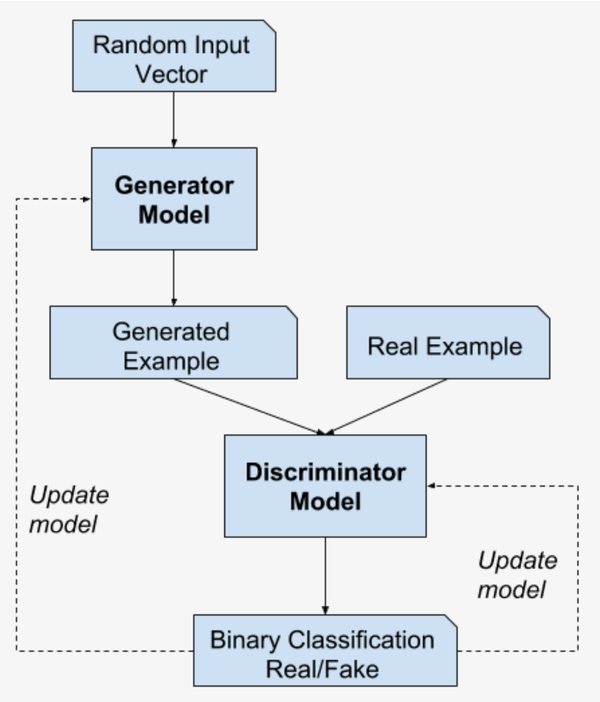
$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Q is a matrix that contains the query (vector representation of one word in the sequence), K are all the keys (vector representations of all the words in the sequence) and V are the values, which are again the vector representations of all the words in the sequence.

Sequence to sequence type model, initially designed with NLP tasks in mind.

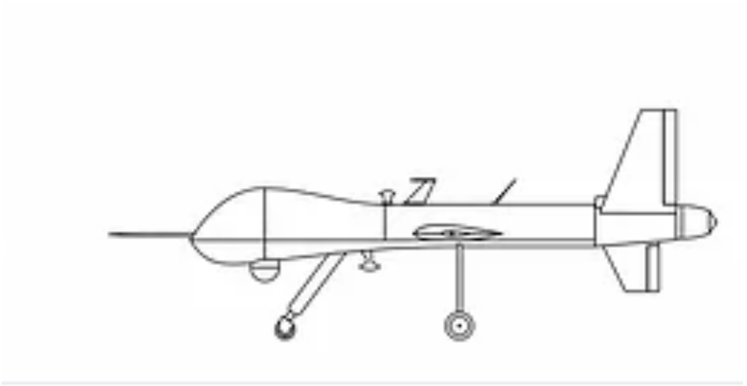
BERT transformer - 110 000 000 trainable parameters
GPT-3 175 billion trainable parameters

Deep learning: Generative Adversarial Networks (I. Goodfellow et al. 2014)



The two models, the generator and discriminator, are trained together. The generator generates a batch of samples, and these, along with real examples from the domain, are provided to the discriminator and classified as real or fake.

The discriminator is then updated to get better at discriminating real and fake samples in the next round, and importantly, the generator is updated based on how well, or not, the generated samples fooled the discriminator.

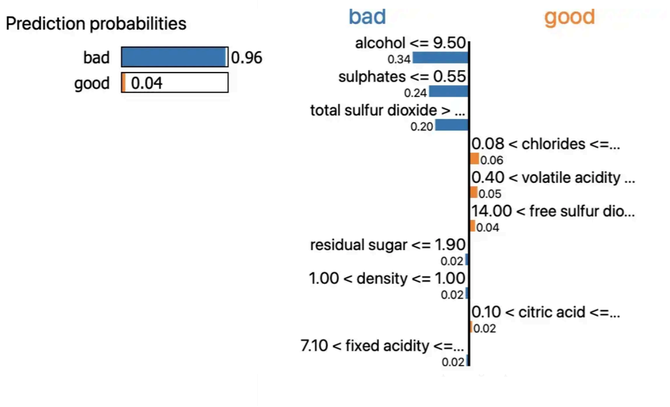


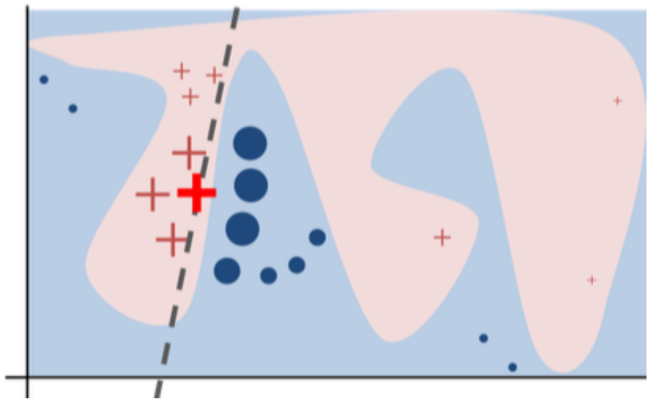
Explainable Machine Learning

Usually referred as explainable artificial intelligence (XAI)
In the case of numerical data provides the contribution of each feature in the decision making process.

Taxonomy:

1. Removal based techniques: Shapley values inspired SHAP package
2. Gradient based techniques.
3. Propagation based techniques.
4. Other methods: self explainable structures, local linear explanations etc., like LIME





Original Image

Interpretable Components

Dank je wel!
Thank you!